U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL OFFICE NOTE 82-7

FURTHER DEVELOPMENT AND TESTING OF AN AUTOMATED SYSTEM TO FORECAST SNOW AMOUNTS

Joseph R. Bocchieri

June 1982

FURTHER DEVELOPMENT AND TESTING OF AN AUTOMATED SYSTEM TO FORECAST SNOW AMOUNTS

Joseph R. Bocchieri

1. INTRODUCTION

A Model Output Statistics (MOS) system for forecasting heavy snow has been operational within the National Weather Service since October 1977 (Bocchieri, 1979; National Weather Service, 1978). Heavy snow is defined as a fall of $\geq$4 inches during a 12-h period at a station. In MOS (Glahn and Lowry, 1972), a statistical relationship is determined between the forecast output of a numerical prediction model (or models) and observed occurrences of a particular weather element. For the operational heavy snow system, we used output from both the Limited-area Fine Mesh (LFM) model (National Weather Service, 1971; Gerrity, 1977) and a finer mesh version of the LFM, called LFM-II (National Weather Service, 1977; Newell and Deaven, 1981), to develop prediction equations that give the conditional probability of heavy snow for the 12-24 h period after 0000 GMT and 1200 GMT. An estimate of the unconditional probability of heavy snow and a categorical forecast are also provided.

We recently did several experiments with the goal of developing an improved operational system which would provide forecasts for other snow amount categories, not just heavy snow, and other forecast projections, in addition to the 12-24 h projection (Bocchieri, 1982a). Based on verification of experimental snow amount forecasts on independent data, we concluded that forecasts for the 12-18, 18-24, and 24-36 h periods weren't good enough for operational implementation of the equations for these periods at this time. On the other hand, we felt that the results for the 12-24 h projection were promising, even though there was some deterioration of the verification scores on the independent data sample. We also found that the experimental forecasts for the heavy snow category for the 12-24 h projection were of comparable accuracy to the present operational forecasts.

To improve the stability of the forecast equations for the 12-24 h projection equations, we derived new equations with a larger developmental sample. This report describes the developments and further testing of these new snow amount forecast equations.

2. DEVELOPMENT OF NEW SNOW AMOUNT FORECAST EQUATIONS

The potential predictors from the LFM and LFM-II models used in the development of the new conditional probability of snow amount [PoSA(S)] equations are listed in Table 1. Model output variables valid for 6-, 12-, 18-, and 24-h projections were included in addition to station elevation and the sine and cosine of the day of year. The predictors were used in both unsmoothed and 9-point, space-smoothed form.

To form the predictand for the PoSA(S) equations, we included only "pure snow" events in our developmental sample. A pure snow event is defined as the occurrence at a station of $\geq$0.1 inches of snow and/or sleet, and no other type of precipitation, during a 12-h period. Therefore, the PoSA(S) equations produce probability forecasts for a site given that a pure snow event occurs.

Actually, we could isolate only quasi-pure snow events in the developmental sample. The basic data were 6-h snowfall amounts at 195 stations in the conterminous United States. In order to isolate the pure snow events we examined the "weather" observations to determine the type of precipitation falling within the period. However, weather observations were available only every third hour. Therefore, since we couldn't be certain that only pure snow fell within the 12-h period, a number of snow events may have been only quasi-pure.

We used the Regression Estimation of Event Probability (REEP) screening technique (Miller, 1964) to develop PoSA(S) forecast equations. This technique objectively selects a subset of effective predictors from a large set of potential predictors to use in multiple linear regression equations. The equations give estimates of the probabilities of occurrence of a given set of binary-type predicands. Snow amount was categorized into three, cumulative, binary-type predictands: ≥ 2, ≥ 4, and ≥ 6 inches[1]. The predictand is called binary because in the developmental phase it was assigned a value of 1 or 0 for a given case depending, respectively, upon whether or not that particular snow amount category occurred. The potential predictors in Table 1 were included in both binary and continuous form. The use of binary predictors helps to account for non-linear relationships which may exist between the predictors and predictands. A good description of the screening procedure can be found in Glahn and Lowry (1972).

PoSA(S) equations were derived for each of several geographic regions which were determined in the following manner. First, we derived PoSA(S) equations for the ≥ 2 and ≥ 4 inch categories for the 12-24 h period from both 0000 GMT and 1200 GMT with data combined from 195 stations--the so-called generalized operator approach. The developmental sample consisted of nine winter seasons, October through March, 1972-73 through 1980-81. We then evaluated the equations on the developmental sample to obtain forecasts for each station. A statistic called the relative probability bias was computed for each station and for each snow amount category by:

$$\text{Relative Probability Bias} = \frac{\overline{\text{PoSA(S)}} - \text{RF}}{\text{RF}} ,$$

where $\overline{\text{PoSA(S)}}$ is the average conditional probability forecast for a particular snow amount category for each station and RF is the relative frequency of that category for each station. We subjectively determined the regions shown in Fig. 1 by grouping stations having similar relative probability bias characteristics; other factors we considered were the density of stations and the climatic frequency of snow amount. We needed to make the regions large to insure that a sufficient number of snow amount cases would be available for equation development. Stations to the south or west of the dashed lines in Fig. 1 were not included because we don't archive snow amount data for stations in the extreme South and in southern California. We do archive snow amount data for stations in northern California, western Oregon, and western

---

[1]For operational purposes, these break points are given in whole inches. However, for developing the equations, we used ≥ 1.5, ≥ 3.5, and ≥ 5.5 inches as the breakpoints, since we archive snow amounts to the nearest tenth of an inch.

Washington; however, the relative probability bias values indicated that these stations shouldn't be included in region 1, and there weren't enough snow cases to allow us to make a region which included only these stations.

PoSA(S) equations were then developed for each region for the 12-24 h projections from 0000 GMT and 1200 GMT for the ≥ 2, ≥ 4, and ≥ 6 inch categories by combining data from all stations within the region. The equations included 12 predictors because experimental results indicate this number to be about optimum (Bocchieri, 1982a).

In Table 2, the LFM predictors are ranked on the basis of a weighted scoring system that emphasizes both frequency and order of selection of predictors by the REEP screening procedure. For this purpose, all predictor projections, smoothings, and binary limits were combined for each type of variable. The screening results for all regions and for both 0000 GMT and 1200 GMT were also combined. The results in Table 2 indicate that, generally, the most important predictor was the LFM precipitation amount forecast followed by the mean relative humidity forecast. Other predictors ranked among the first seven included measures of circulation intensity such as 700-mb vertical velocity, 700-mb east-west wind component, 850-mb divergence, 850-mb relative vorticity, and 850-mb east-west wind component. A similar ranking given in Bocchieri (1982a) based on screening results for a smaller developmental sample also showed that LFM forecasts of precipitation amount and mean relative humidity were, generally, the most important predictors.

Table 3 shows the total reduction of variance for the PoSA(S) equations for each region. The relative frequency for each snow amount category and the total number of pure snow cases are also given. As expected, the reduction of variance was generally lower over the western, more mountainous portion of the United States (regions 1 and 2); it's generally more difficult to predict snow amounts in mountainous areas because of local, orographic effects. Also, as expected, the reduction of variance was generally lower for the higher, less frequent snow amount categories. However, in regions 1, 3, 4, and 5 the reduction of variance for the ≥ 4 and ≥ 2 inch categories was about the same even though the later category occurred much more frequently. A possible reason is that in these regions weather systems producing at least 4 inches of snow are generally more organized, occur on a larger scale, and are easier to predict than weather systems producing lesser snow amounts.

To produce unconditional probability of snow amount (PoSA) forecasts, we simply multiplied the conditional probability forecast, PoSA(S), for each snow amount category, by the probability of precipitation (PoP) (Lowry and Glahn, 1976; National Weather Service, 1980) for the corresponding 12-h period and the average conditional probability of frozen precipitation ($\overline{\text{PoF}}$) for the same 12-h period.[2] To obtain the average $\overline{\text{PoF}}$, we averaged the 12-, 18-, and 24-h PoF forecasts; in this scheme, the 18-h forecast was weighted twice as much as the 12- and 24-h forecasts. The PoF forecasts were obtained from the new probability of precipitation type system which will be implemented within the National Weather Service in the fall of 1982 (Bocchieri and Maglaras, 1982b). For

---

[2] In a true mathematical sense, this method doesn't give the unconditional probability of snow amount for 12-h periods in which mixed precipitation occurs, since PoSA(S) equations were developed with pure snow cases only.

instance, the unconditional probability of the ≥ 2 inch category was estimated by:

$$PoSA(\geq 2 \text{ inch}) \approx PoSA(S)(\geq 2 \text{ inch}) \times PoP \times \overline{PoF}. \qquad (2)$$

3. VERIFICATION OF SNOW AMOUNT FORECASTS

We first computed the bias, threat score, and post-agreement[3] for categorical snow amount forecasts from the new system on independent data and compared these scores to those for the developmental sample. The independent sample consisted of data from 181 conterminous United States stations for the period October 1981 through January 1982.

In order to make categorical snow amount forecasts from the probability forecasts, we developed threshold probability values for each snow amount category, each region, and for both 0000 GMT and 1200 GMT. This was done by maximizing the threat score for each category while restricting the categorical bias to ≤ 1.30. This method is explained in more detail in Bocchieri (1982a).

Table 4 shows the probability threshold values, bias, threat score, and post-agreement for categorical snow amount forecasts for the 12-24 h period from 0000 GMT and 1200 GMT for each region given in Fig. 1. The sample consisted of developmental data from nine winter seasons, 1972-73 through 1980-81. The results indicate the following.

1. The lower probability threshold values were generally associated with the higher, less frequent, snow amount categories. The reason is that, as expected, the probability forecasts from the statistical equations were lower, in the mean, for the rarer events. As stated previously, the threshold values were chosen to maximize the threat score while restricting the bias to ≤ 1.30.

2. For regions 2 through 5, the threat scores ranged from .27 to .34 for ≥ 2 inches, from .18 to .24 for ≥ 4 inches, and from .12 to .25 for ≥ 6 inches. In contrast, the threat scores for region 1 were not as good: near .20 for ≥ 2 inches, between .11 and .14 for ≥ 4 inches, and between .04 and .12 for ≥ 6 inches.

3. The post-agreements for regions 2 through 5 ranged from 38.7% to 48.7% for ≥ 2 inches, 30.2% to 36.9% for ≥ 4 inches, and from 23.9% to 38.9% for ≥ 6 inches. As was the case for the threat scores, the post-agreements for region 1 were not as good; near 32% for ≥ 2 inches, between 19.7% and 25.0% for ≥ 4 inches, and between 8.7% and 23.7% for ≥ 6 inches.

Table 5 shows verification scores for categorical snow amount forecasts for both the developmental and independent data samples. For the purpose of this verification, data were combined from 181 conterminous United States stations. Scores are shown for 0000 GMT, 1200 GMT, and for 0000 GMT and 1200 GMT combined. The results can be summarized as follows.

---

[3]The post-agreement = A/B, the threat score = A/(B+C-A), and the bias = B/C, where A, B, and C are the number of correct forecasts, the total number of forecasts, and the number of observations of the event, respectively.

1. For the 0000 GMT cycle, the threat score and post-agreement were generally better for the independent sample as compared to the developmental sample; the bias was better for ≥ 4 inches but worse for ≥ 2 and ≥ 6 inches.

2. For the 1200 GMT cycle, all the scores were worse on the independent sample, especially for the ≥ 6 inch category.

3. For 0000 GMT and 1200 GMT combined, the threat score and post-agreement for the independent sample were generally similar to these scores for the developmental sample, except there was some decrease in the threat score for the ≥ 6 inch category. The bias was better for ≥ 2 inches but worse for ≥ 4 and ≥ 6 inches.

Based on the results for 0000 GMT and 1200 GMT combined, we concluded that the new snow amount forecast equations are suitable for operational implementation. In Table 5, the instability of the scores on the independent sample for the individual cycle times, as compared to the general stability of the scores for 0000 GMT and 1200 GMT combined, was likely due to the relatively small sample size for each cycle time. We felt that the results for both cycle times combined were more representative of the accuracy of the new system because of the larger sample size.

We also examined the reliability of the unconditional probability forecasts from the new system on the independent data sample. Reliable probability forecasts are such that for all of the forecasts of 20%, say, for a particular category the relative frequency of that category is close to 20%. Figs. 2, 3, and 4 show the reliability for the ≥ 2, ≥ 4, and ≥ 6 inch categories, respectively; for this purpose, data were combined from 181 stations and from 0000 GMT and 1200 GMT. For ≥ 2 inches (Fig. 2), the results indicate that the probability forecasts were generally quite reliable. There was some tendency for overforecasting for probability forecasts >75%, but the number of cases for this range of probabilities was small. The reliability for the probability forecasts for ≥ 4 inches (Fig.3) was not as good. The probability forecasts <35% were reliable, but there was some tendency to overestimate the frequency of occurrence for probabilities >35%. Note the small number of cases for probabilities >55%. For >6 inches (Fig. 4), the probability forecasts were generally reliable except for some tendency to overestimate the frequency of occurrence for probabilities >25%. Note that there were no cases of probability forecasts ≥ 40%.

Figs. 5 through 8 show sample MOS categorical snow amount forecasts for the ≥ 2, ≥ 4, and ≥ 6 inch categories for two cases during January 1982 (independent data). One case is representative of well organized, synoptic-scale systems for which MOS can generally be expected to perform well. The other case is representative of snowfalls associated with mesoscale lake effects and local topography; MOS forecasts are generally poor for these situations.

Figs. 5 through 7 show snow amount forecasts and observations associated with an intense low pressure system which developed over the central Rocky Mountains on January 21, moved to the Central Plains on January 22, and continued to intensify as it moved to the upper Great Lakes on January 23. Also, a low developed near the mid-Atlantic Coast on January 23 on the warm front associated with the intense storm over the Great Lakes. This coastal low also intensified as it moved to Nova Scotia on January 24. Note in Fig. 5 that while the forecasts for the upper Mid-West were generally good, MOS overforecasted snow amounts from the central Appalachians to the lower Great Lakes area. For the 12-h period ending 1200 GMT January 23 (Fig. 6), MOS generally underestimated snow amounts over the upper Mid-West, but again snow amounts were overforecasted in the Great Lakes area. The MOS forecasts for the period ending 0000 GMT January 24 (Fig. 7) were generally good; the heavy amounts over interior New England were associated with the intensifying coastal low.

Fig. 8 is a typical case of poor MOS forecasts of snow amounts associated with lake-effects. On January 9, a cold front moved through the Great Lakes area followed by a northwest flow of arctic air. In Fig. 8 note that MOS underforecasted snow amounts over the eastern shore of Lakes Michigan and Superior and western New York near Lakes Erie and Ontario.

4. SUMMARY

MOS forecast equations for ≥ 2, ≥ 4, and ≥ 6 inch snow amount categories were developed and tested for the 12-24 h projection after both 0000 GMT and 1200 GMT. Previous experimental results indicated that MOS snow amount forecasts for >24 hours were not accurate enough to consider operational implementation at this time. Verification of the new snow amount equations indicates that scores were generally stable on independent data. Therefore, we plan to implement a new MOS snow amount forecast system in September 1982. This new system should be more useful than the present operational system which provides forecasts only for the ≥ 4 inch category.

5. ACKNOWLEDGEMENT

The author wishes to thank the many members of the Techniques Development Laboratory who developed and maintain the MOS system.

6. REFERENCES

Bocchieri, J. R., 1979: The use of LFM output for automated prediction of heavy snow. Preprints Fourth Conference on Numerical Weather Prediction, Silver Spring, Amer. Meteor. Soc., 77-81.

______, 1982a: Recent experiments in the use of model output statistics for forecasting snow amounts. TDL Office Note 82-2, National Weather Service, NOAA, U.S. Department of Commerce, 7 pp.

______, and G. J. Maglaras, 1982b: Recent improvements in an automated system for forecasting precipitation type. TDL Office Note 82-3, National Weather Service, NOAA, U.S. Department of Commerce, 37 pp.

Gerrity, J. F., Jr., 1977: The LFM model--1976: a documentation. NOAA Technical Memorandum NWS NMC-60, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 68 pp.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteor., 11, 1203-1211.

Lowry, D. A., and H. R. Glahn, 1976: An operational model for forecasting probability of precipitation--PEATMOS PoP. Mon. Wea. Rev., 104, 221-231.

Miller, R. G., 1964: Regression estimation of event probabilities. Tech. Rep. No. 1, Contract CWB-10704, The Travelers Research Center, Inc. Hartford, Conn., 153 pp. [NTIS AD 602037].

National Weather Service, 1971: The limited-area fine mesh (LFM) model. NWS Technical Procedures Bulletin No. 67, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 11 pp.

______, 1977: High resolusion LFM (LFM-II). NWS Technical Procedures Bulletin No. 206, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 6 pp.

______, 1978: The use of model output statistics for predicting the probability of heavy snow. NWS Technical Procedures Bulletin No. 246, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 10 pp.

______, 1980: The use of model output statistics for predicting probability of precipitation. NWS Technical Procedures Bulletin No. 289, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 13 pp.

Newell, J. E., and D. G. Deaven, 1981: The LFM-II model--1980. NOAA Technical Memorandum NWS NMC-66, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 20 pp.

Table 1. The potential predictors used to develop the conditional probability of snow amount equations.

Sea-level pressure tendency	700-mb east-west wind component
850-mb height	700-mb north-south wind component
850-mb height tendency	500-mb height tendency
850-mb relative vorticity	500-mb relative vorticity
850-mb temperature advection	500-mb vorticity advection
850-mb divergence	Advection of 1000-500 mb thickness by the 700-mb geostrophic wind
850-mb moisture convergence	500-mb east-west wind component
850-mb east-west wind component	500-mb north-south wind component
850-mb north-south wind component	Mean relative humidity (sfc.-500 mb)
700-mb relative vorticity	Precipitation amount
700-mb divergence	Station elevation
700-mb moisture convergence	Sine of the day of year
700-mb temperature advection	Cosine of the day of year
700-mb vertical velocity	

Table 2. The rank of LFM predictors in the conditional probability of snow amount forecast equations on the basis of frequency and order of selection by the REEP screening procedure. For this analysis, screening results for all regions and for 0000 GMT and 1200 GMT were combined.

Rank	Predictor Type
1	Precipitation amount
2	Mean relative humidity (sfc.-500 mb)
3	700-mb veritical velocity
4	700-mb east-west wind component
5	850-mb divergence
6	850-mb relative vorticity
7	850-mb east-west wind component
8	Precipitable water (sfc.-500 mb)
9	850-mb moisture convergence
10	700-mb moisture convergence
11	1000-500 mb thickness advection
12	700-mb north-south wind component

Table 3. The reduction of variance for REEP conditional probability of snow amount equations for the 12-24 h projection from 0000 GMT for each of the five regions shown in Fig. 1. The developmental data were from nine winter seasons (1972-73 through 1980-81). The relative frequency (%) of each snow amount category is also shown in parentheses.

Region	Total Reduction of Variance (%)			Total Number of Snow Cases
	Snow Amount Category			
	$\geq$2 inches	$\geq$4 inches	$\geq$6 inches	
1--38 stations	11.6 (23.0)	13.1 (5.0)	7.1 (1.4)	2638
2--10 stations	17.2 (34.2)	12.9 (13.3)	11.7 (4.1)	1107
3--40 stations	21.7 (20.4)	20.1 (4.7)	15.8 (1.2)	4621
4--24 stations	19.4 (26.9)	19.1 (8.0)	18.7 (2.5)	4518
5--72 stations	22.2 (33.7)	22.0 (12.6)	15.2 (5.0)	1962

Table 4. The probability threshold values (PT) (values in percent), bias, threat score (TS), and post-agreement (PA) (values in percent) for categorical snow amount forecasts for each region shown in Fig. 1. The sample consists of developmental data from the winter season of 1972-73 through 1980-81. The projection is the 12-24 h period from 0000 GMT and 1200 GMT.

Region	Snow Amount Category											
	≥2 inches				≥4 inches				≥6 inches			
	PT	Bias	TS	PA	PT	Bias	TS	PA	PT	Bias	TS	PA
0000 GMT												
1	18.8	1.08	.20	31.5	11.8	1.01	.14	25.0	10.8	.79	.12	23.7
2	26.0	1.15	.27	39.3	21.0	1.04	.18	30.2	16.2	.83	.18	33.3
3	22.8	1.28	.28	38.7	21.0	1.10	.20	32.2	19.8	.90	.16	28.3
4	25.0	1.07	.31	45.4	23.0	1.09	.24	36.6	22.9	1.03	.25	38.9
5	27.8	1.03	.29	44.9	23.0	1.15	.19	30.3	20.4	.77	.12	23.9
1200 GMT												
1	19.0	1.07	.20	32.2	10.6	1.05	.11	19.7	6.6	.70	.04	8.7
2	22.0	1.28	.28	39.6	17.2	1.23	.22	33.0	18.4	.90	.17	30.0
3	27.6	1.04	.30	45.3	20.0	1.12	.21	32.5	20.6	.91	.14	26.2
4	22.8	1.09	.31	44.9	22.8	1.14	.24	36.9	24.2	.91	.22	38.4
5	28.6	1.09	.34	48.7	21.6	1.26	.23	33.7	23.2	.92	.16	28.4

Table 5. The bias, threat score, post-agreement, and number of cases for categorical snow amount forecasts for 181 stations combined. The developmental sample (D) consisted of nine winter seasons, 1972-73 through 1980-81. The independent sample (I) period was from October 1981 through January 1982. The results are shown for 0000 GMT, 1200 GMT, and both cycles combined.

Snow Amount Category (inches)	Bias		Threat Score		Post-Agreement (%)		Number of Cases	
	D	I	D	I	D	I	D	I
0000 GMT								
≥2	1.12	1.17	.27	.28	40.7	40.0	4716	421
≥4	1.09	1.02	.20	.24	31.9	38.3	1300	125
≥6	.88	.72	.17	.19	31.1	38.2	399	47
1200 GMT								
≥2	1.09	.89	.29	.23	43.1	39.7	4987	400
≥4	1.16	.67	.21	.12	32.7	27.1	1298	104
≥6	.89	.43	.17	.03	30.6	10.5	395	44
0000 GMT and 1200 GMT								
≥2	1.11	1.03	.28	.25	41.9	39.9	9703	821
≥4	1.13	.86	.21	.19	32.3	34.3	2598	229
≥6	.89	.58	.17	.12	30.8	28.3	794	91

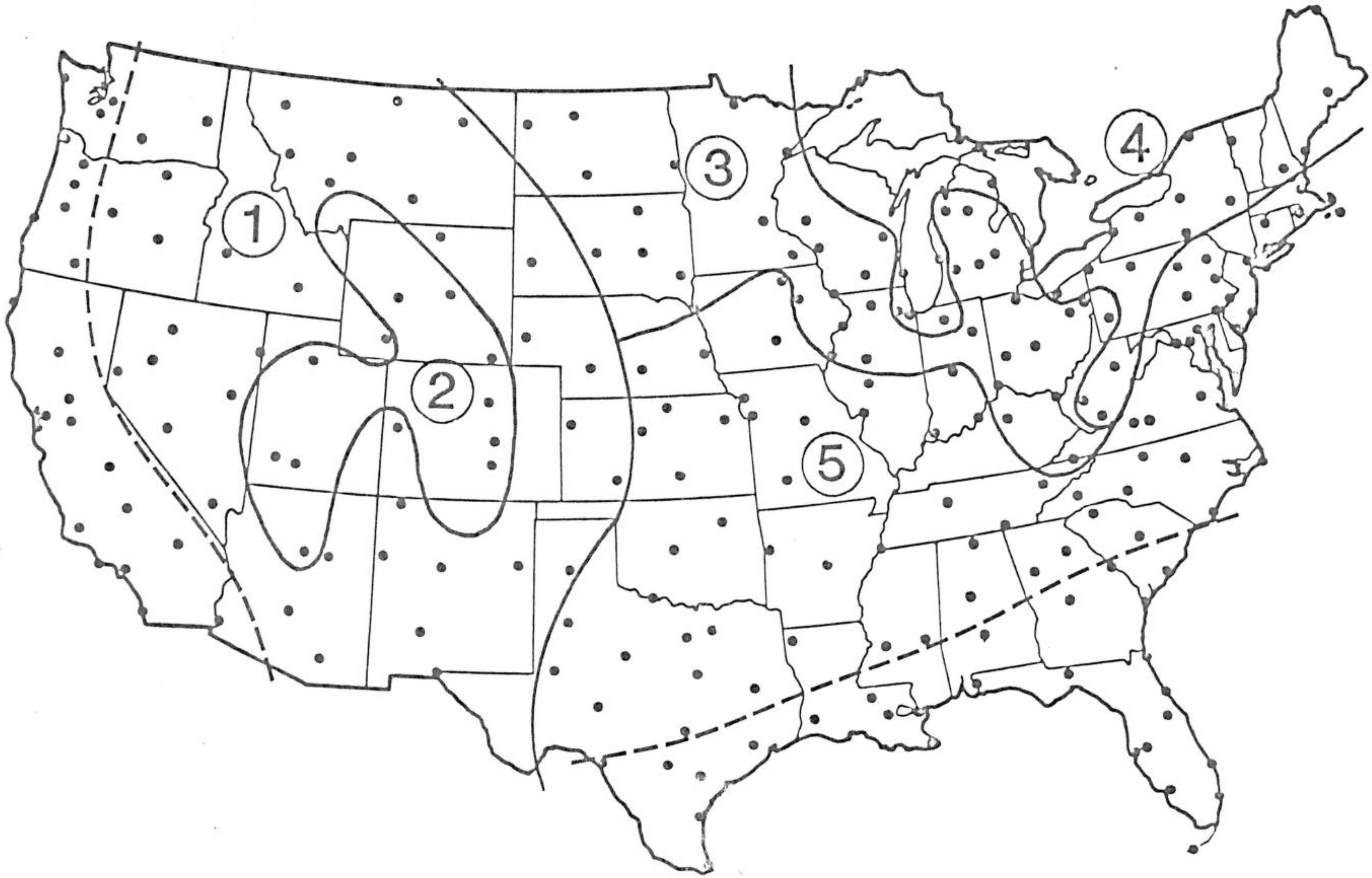

Figure 1. The five regions used in the development of conditional probability of snow amount equations. Stations to the south or west of the dashed lines were not included in the development.

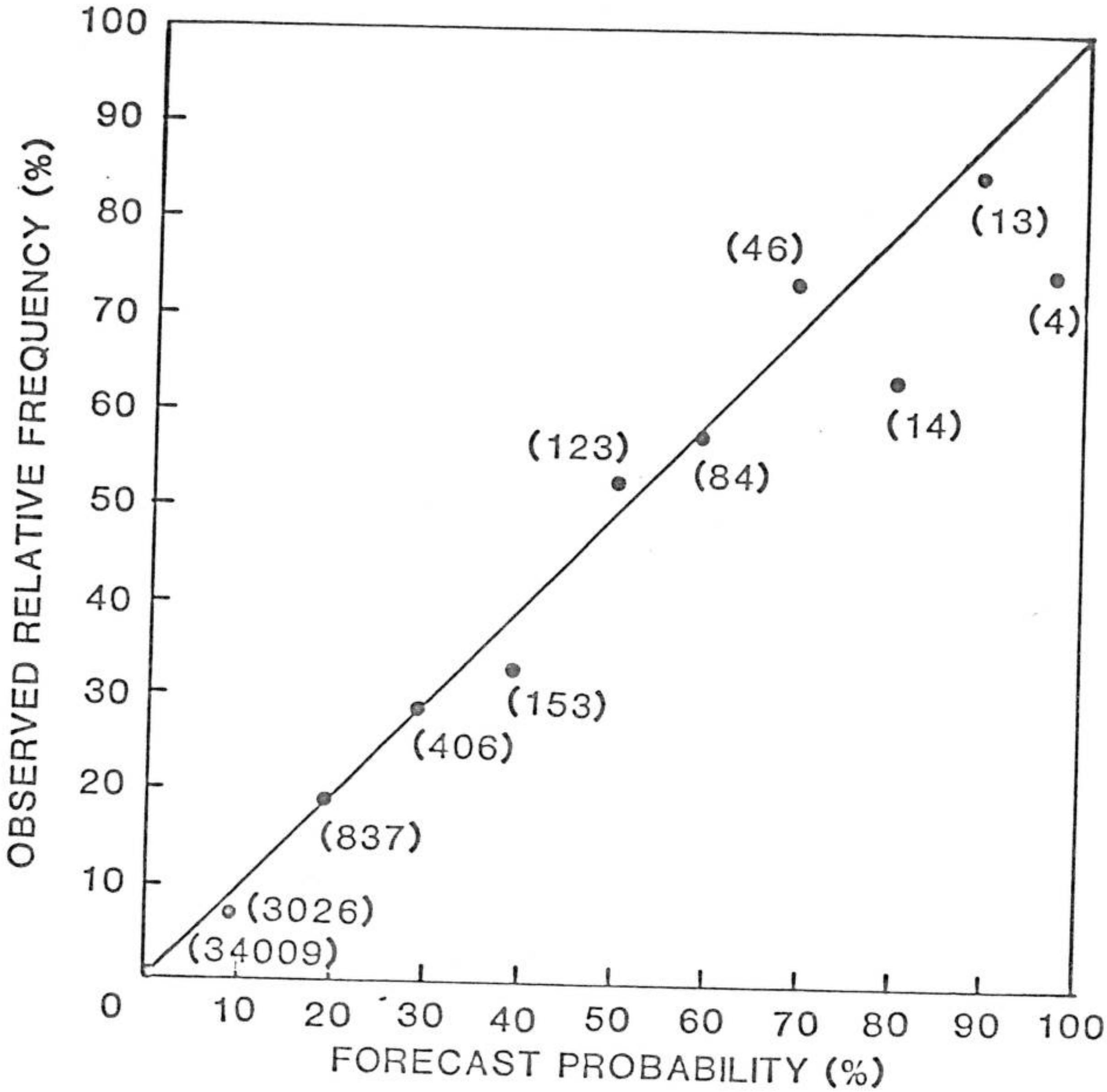

Figure 2. The reliability of unconditional probability forecasts for the ≥ 2 inch snow amount category for the 12-24 h projection. Independent data were combined from 181 stations and from 0000 GMT and 1200 GMT for the period October 1981-January 1982. The number of cases for each data point is shown in parentheses. The line denotes perfect reliability.

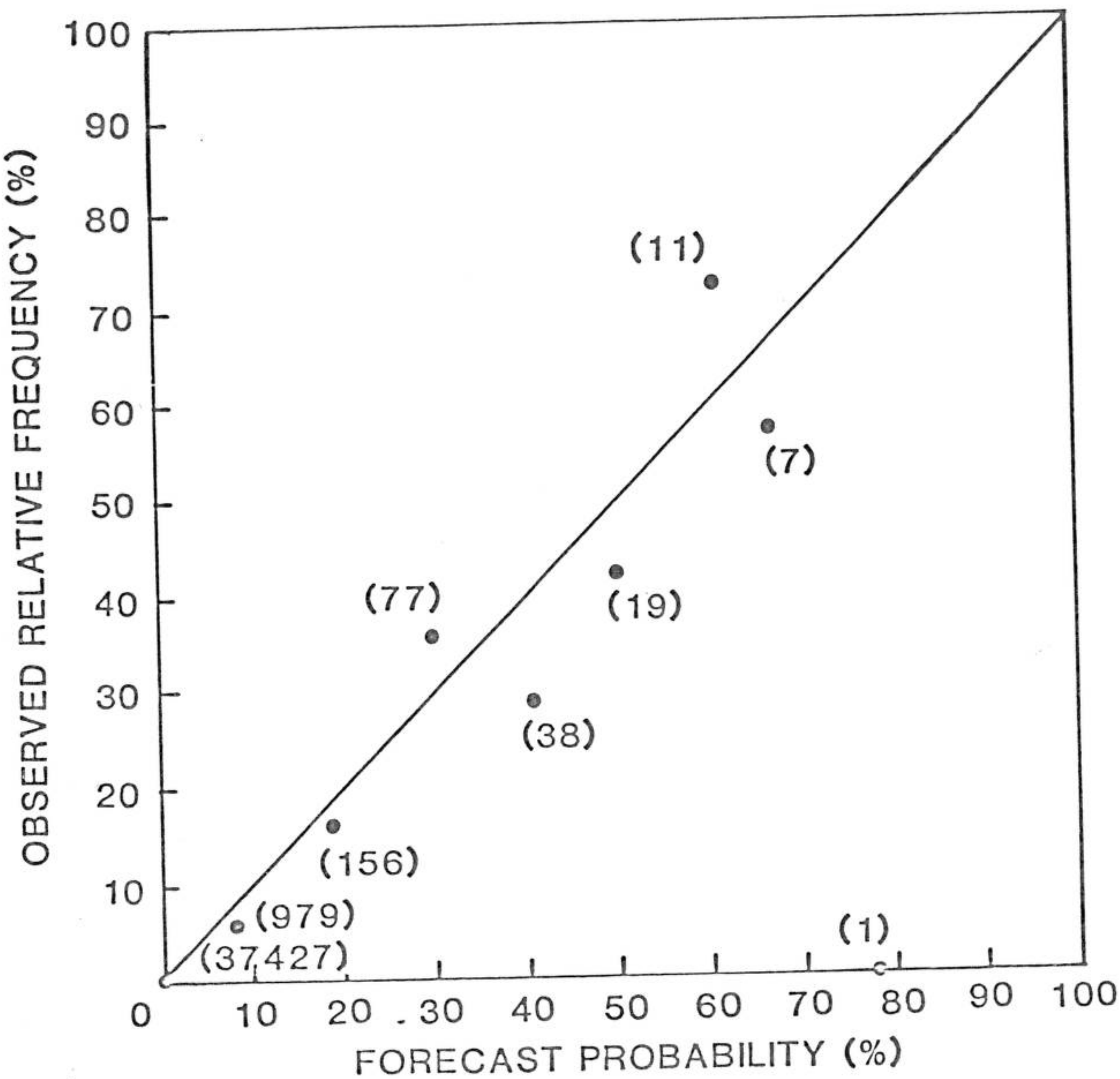

Figure 3. Same as Fig. 2 except for the ≥4 inch category.

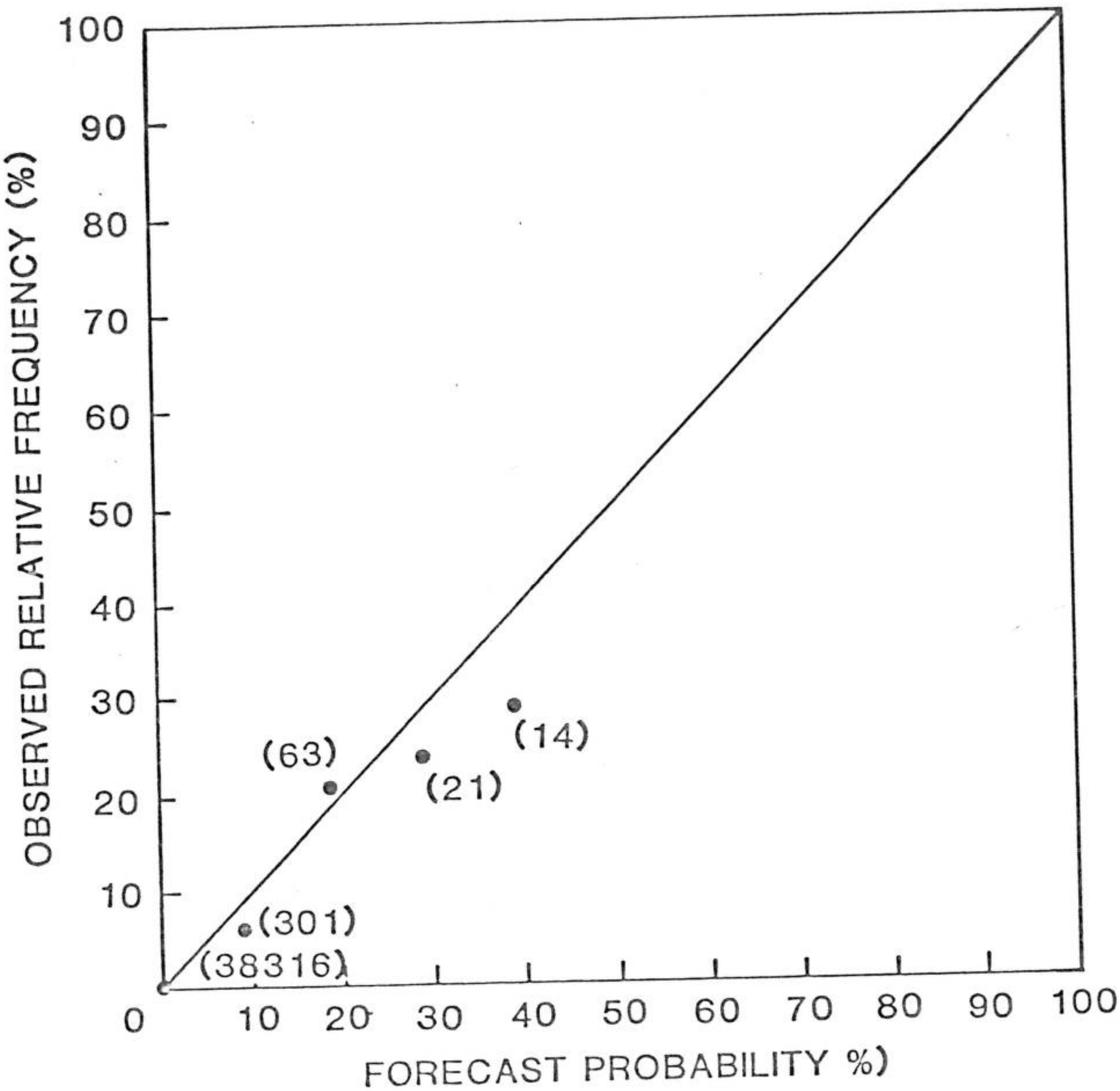

Figure 4. Same as Fig. 2 except for the ≥ 6 inch category.

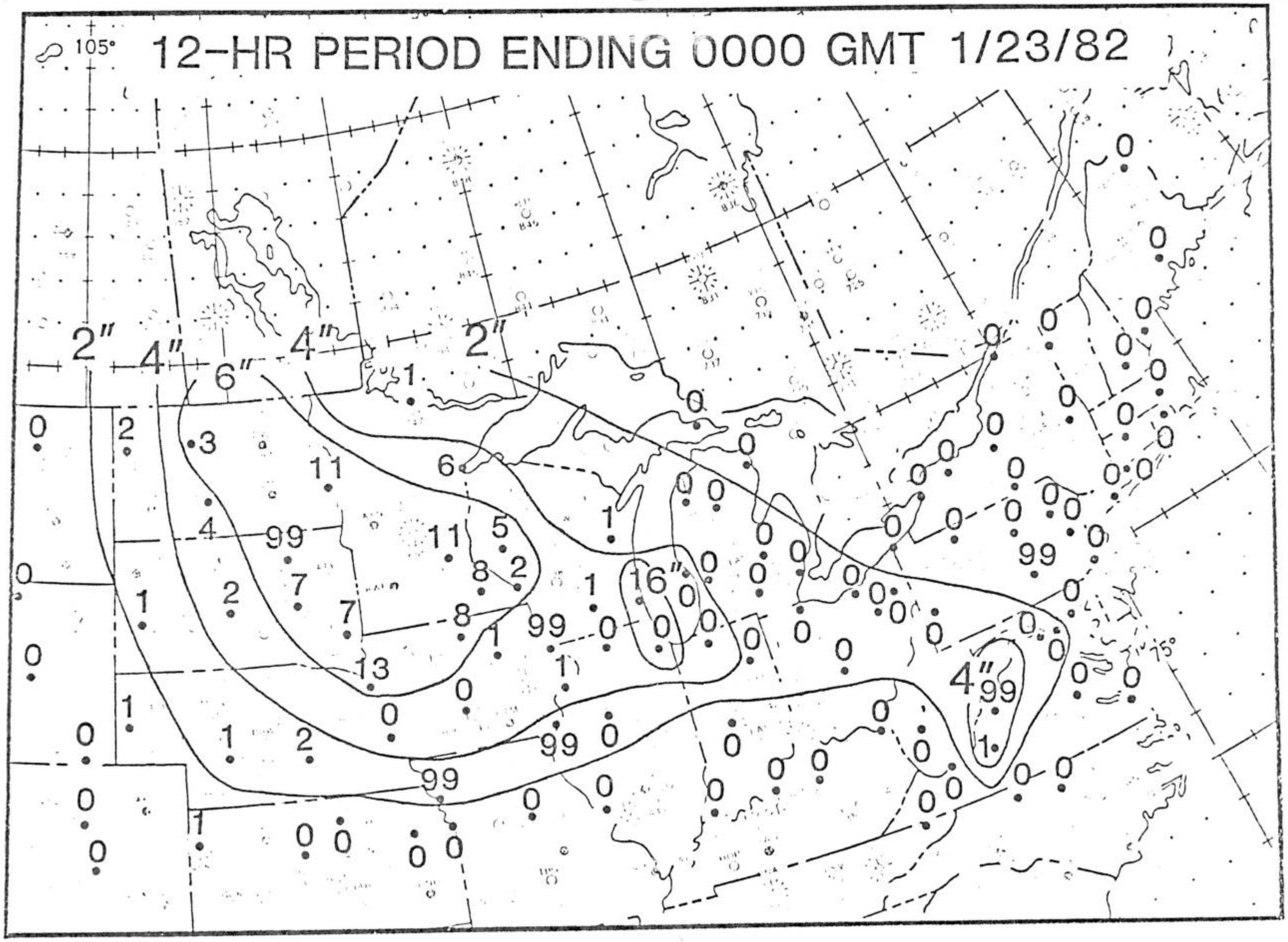

Figure 5. Sample 12-24 h MOS snow amount forecasts for the period ending 0000 GMT, January 23, 1982. Isolines for the ≥ 2, ≥ 4, and ≥ 6 inch snow amount forecast categories are shown. Observed snow amounts (to the nearest inch) for the 12-h period are plotted for MOS stations (99 indicates missing).

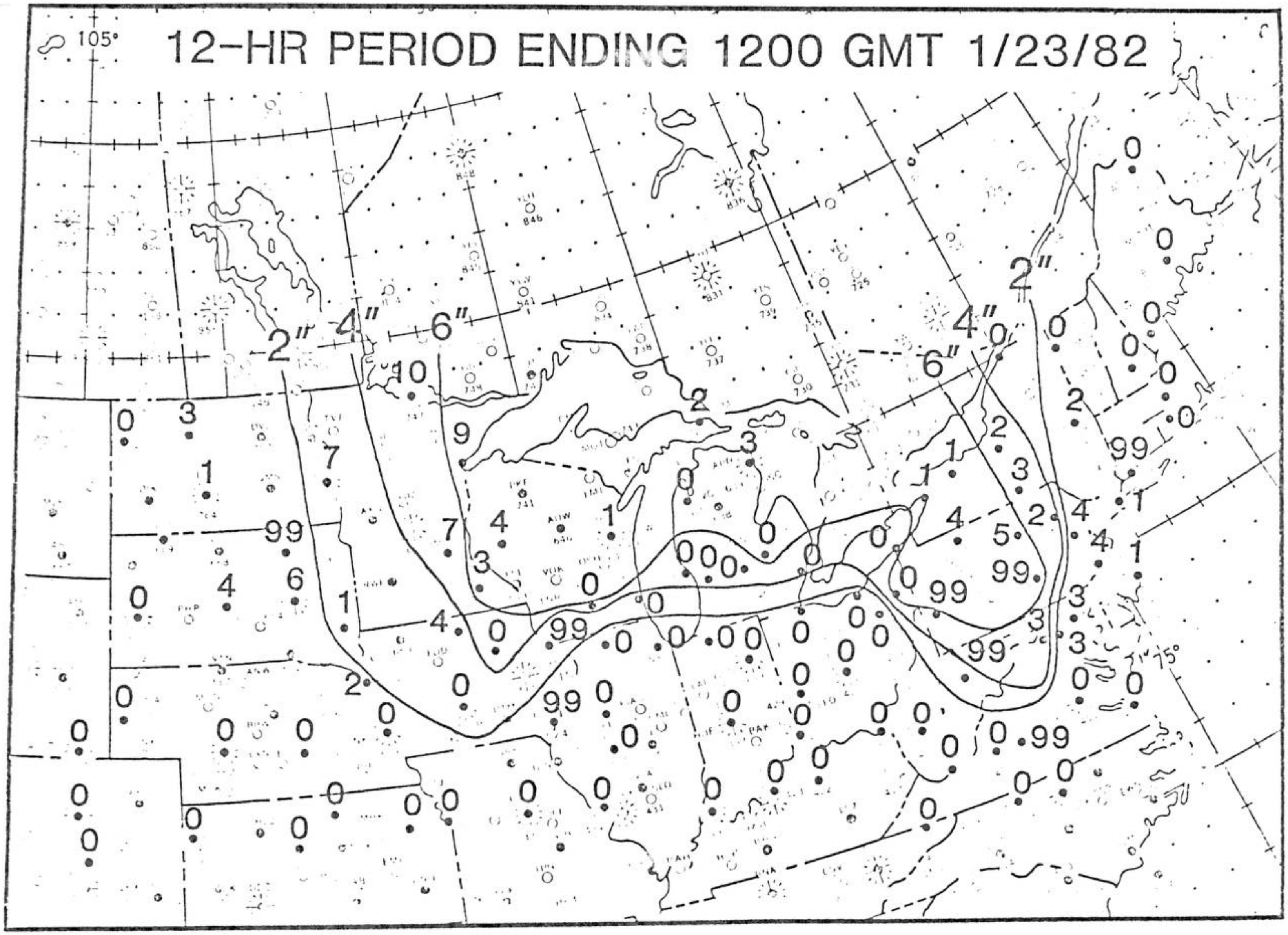

Figure 6. Same as Fig. 5 except the 12-24 h forecasts are valid for the period ending 1200 GMT, January 23, 1982.

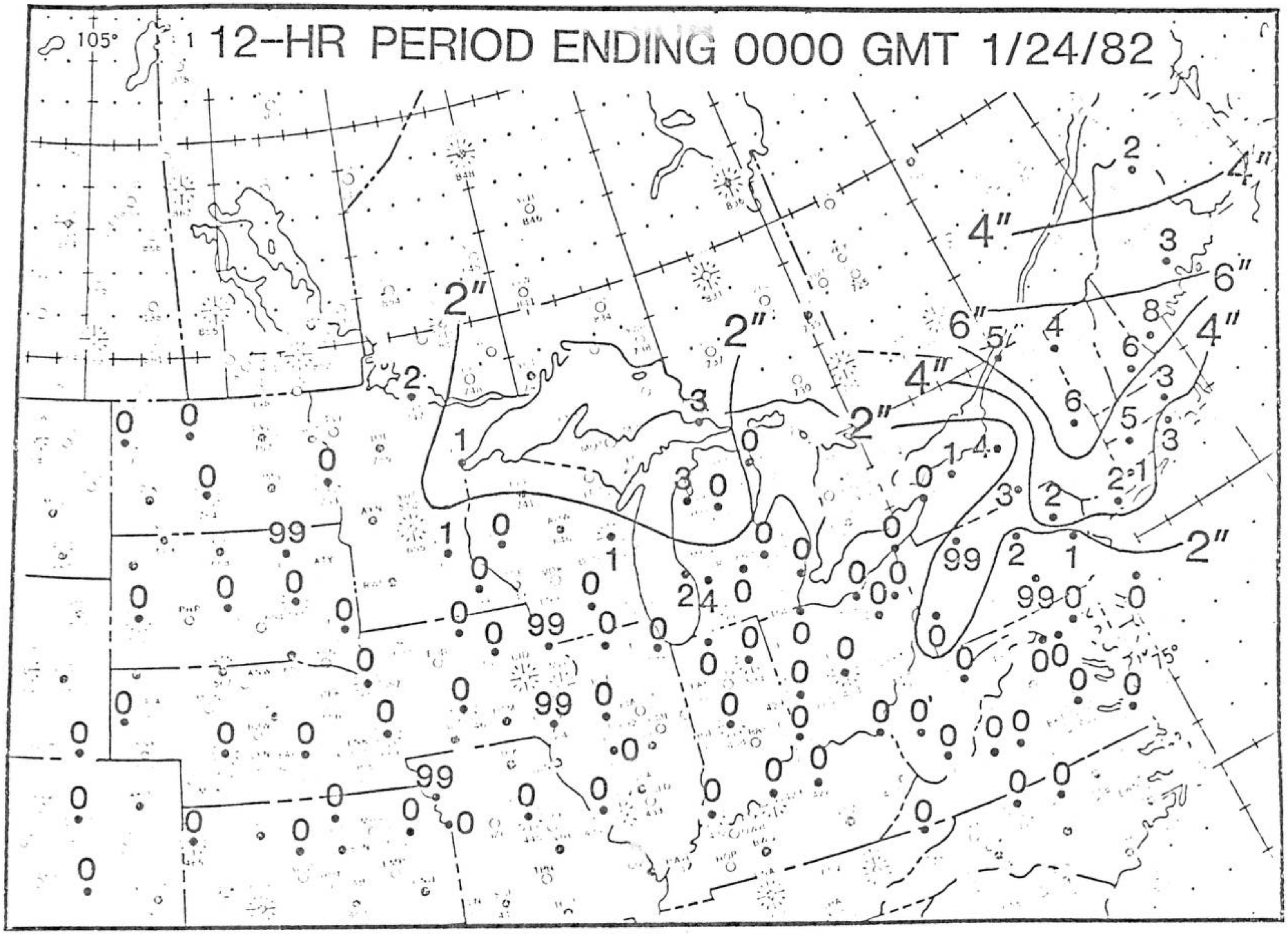

Figure 7. Same as Fig. 5 except 12-24 h forecasts are valid for period ending 0000 GMT January 24, 1982.

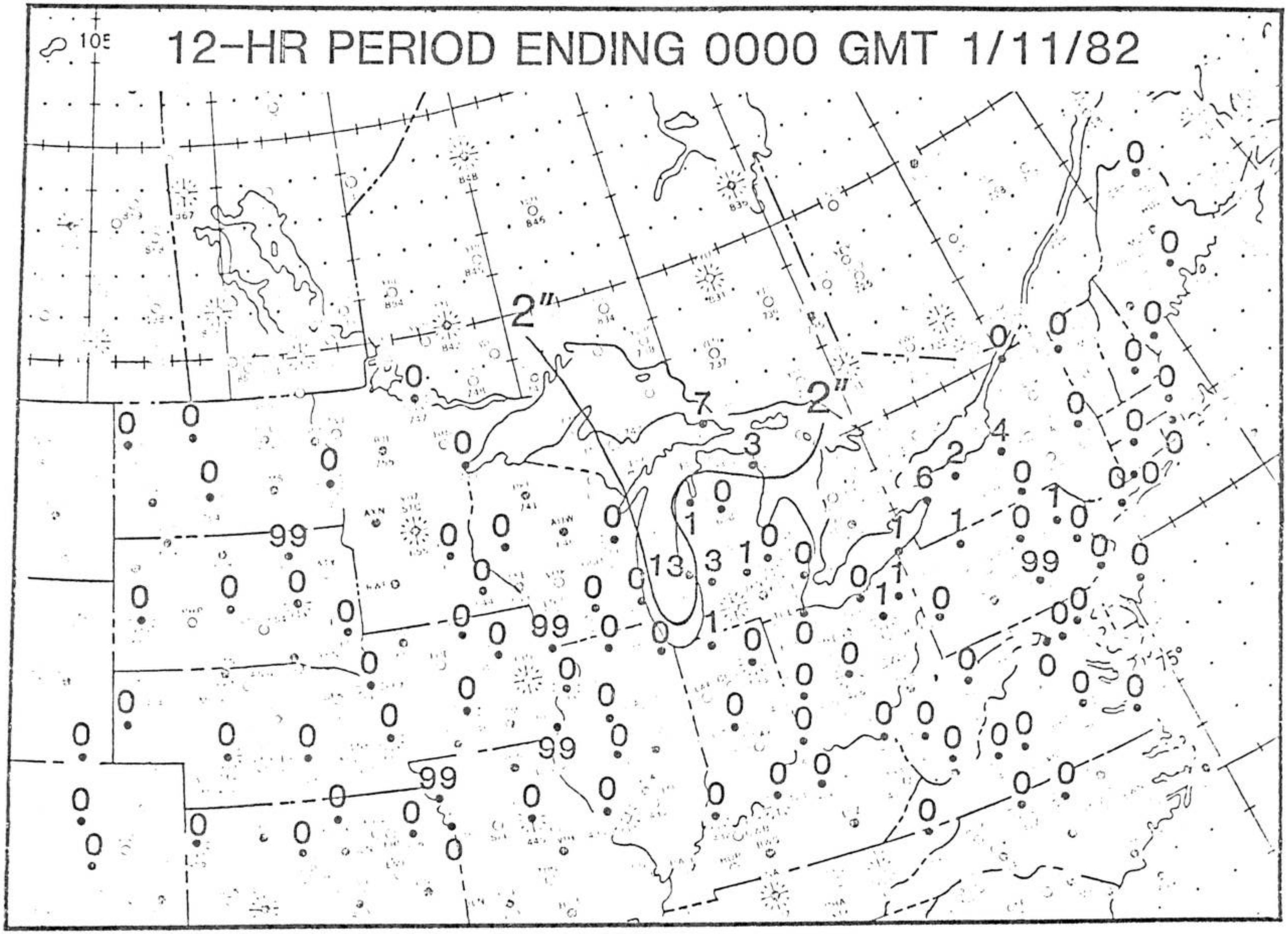

Figure 8. Same as Fig. 5 except the 12-24 h forecasts are valid for the period ending 0000 GMT, January 11, 1982.